Computational Statistics
Part 1
Simulating Statistical Models

M.Sc. Mathematical Engineering @ Politecnico di Milano

Class notes, A.A. 2020-21 (1^{st} Edition)

Andrea Manzoni

One of the topics of this course is the study of statistical models using computer simulations. Here we use the term *statistical models* to mean any mathematical models which include a random component. Our interest in this chapter is in *simulation* of the random component of these models. The basic building block of such simulations is the ability to generate random numbers on a computer (Chapters 1-2). Generation of random numbers, or more general random objects, on a computer is complicated by the fact that computer programs are inherently deterministic: while the output of computer program may look random, it is obtained by executing the steps of some algorithm and thus is totally predictable. The problem of generating random numbers is then split into two distinct subproblems:

1. generating any randomness at all, concentrating on the simple case of generating independent random numbers, uniformly distributed on $[0, 1]$ (Chapter 1);

2. generating random numbers from different distributions, using independent, uniformly distributed random numbers as a basis (Chapter 2).

Afterwards, we introduce the notion of random field, as a general tool to parametrize random *inputs* of models involving differential problems that will be addressed in the subsequent Parts. In particular, we show how the Karhunen-Loeve expansion, based on the eigendecomposition of a suitable compact operator, is of key importance to approximate random fields by means of finite-dimensional objects, thus allowing us to parametrize random fields, and making the ability to sample from statistical distributions even more important (Chapter 3).

Being able to generate random numbers, and sample from statistical distributions, is a building block of Monte Carlo algorithms that are addressed in Chapter 4. In particular, we introduce the Monte Carlo method, showing its properties, its connection with known quadrature formulas, after recalling some asymptotic results. In order to improve its convergence rate, we introduce in Chapter 5 a series of variance reduction techniques, such as antithetic variables, importance sampling, control variates, stratified sampling, latin hypercube sampling. Finally, we address in Chapter 6 some ideas related with quasi Monte Carlo methods, which rely on suitable alternatives to (pseudo) random sampling such as low-discrepancy points sets or sequences, and are capable of reaching a better convergence rate than Monte Carlo methods.

Part 2
Sensitivity Analysis
& Forward Uncertainty Quantification

Computational Statistics - M.Sc. Mathematical Engineering - Politecnico di Milano, A.A. 2020-21

Andrea Manzoni

While building and using numerical simulation models, Sensitivity Analysis (SA) methods are invaluable tools. They allow to study how the uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input. It may be used to determine the most contributing input variables to an output behavior as the non-influential inputs, or ascertain some interaction effects within the model. The objectives of SA are numerous; one can mention model verification and understanding, model simplifying and factor prioritization. Finally, the SA is an aid in the validation of a computer code, guidance research efforts, or the justification in terms of system design safety. In this chapter we review some general strategies to perform SA, following the book by Saltelli et al. [27]. We first review the most common local methods, then turning to the most relevant (global) variance-based methods, introducing first order effects and total effects and the way they can be estimated in practice. We quickly describe elementary effects methods, and turn to the Then, following [9, Chapters 31–39], we will review variational methods, importance measures, derivative-based global sensitivity measures, moment-Independent and reliability-based importance measures, to conclude by considering metamodel-based sensitivity analysis through polynomial chaos expansions and Gaussian processes.

Then, we will turn to forward Uncertainty Quantification (UQ), whose goal is to derive information about the uncertainties in system outputs of interest, given information about the uncertainties in the system (random) inputs. The goal is to obtain the probability density function (PDF) of some output quantity of interest (QoI) given the probability distribution of the input, or evaluate moments, correlation functions, confidence regions or quantiles or, ultimately, to approximate from a numerical standpoint the solution of the problem, seen as a function of the random inputs.

We first address the case of the evaluation of some output QoI, focusing on the case of Monte Carlo and multi level Monte Carlo sampling. Then, we will address some basic notions about stochastic finite element methods, which refer to an extensive class of algorithms for the approximate solution of partial differential equations having random input data, for which spatial discretization is carried out by a finite element method.